A Novel Real-Time Bidding Model Based on Video Processing Technology

Jiesheng Zhang1, Chenyan Feng1, Vincent Chang1, Jia Tan1 and Tian Qin1

University of Michigan – Shanghai Jiao Tong University Joint Institute,
Shanghai Jiao Tong University, Shanghai, China1
ji-designexpo@sjtu.edu.cn1
vincent.chang@sjtu.edu.cn1
Corresponding Author: ji-designexpo@sjtu.edu.cn

Abstract

The aim of this paper is to introduce a novel real-time bidding (RTB) model based on video processing technology and to demonstrate how it can promote business innovation. The proposed RTB model highlighted automatic video processing, manual correction mechanism, and specific customer preferences. To make the model technically feasible, a video processing technology was further developed. This technology was built around four algorithms, including scene-splitting, object recognition, box selection and object tracking, one manual correction mechanism, and one MATLAB Graphical User Interface (GUI). In addition, customers’ preferences are provided to Data-Management Platforms (DMPs), which serves as a price reference for advertisers. With the proposed model, customers are intimately connected to whatever they see and want across all video platforms. At the same time, advertising becomes more powerful in that customer experience is personalized and purchase decision is driven faster.

Keywords: Real-time bidding (RTB), online video advertising, video processing, object recognition, object tracking, and personalized customer experience.

1. Introduction

With the rapid development of the internet, the market of online advertising has been growing explosively. According to the iResearch (2010), the global market size of the online advertising is expected to expand from $68.7 billion in 2011 to $96.8 billion in 2014, with a compound annual growth rate (CAGR) of over 10%. In addition, according to the Interactive Advertising Bureau and PricewaterhouseCoopers LLP (2014), the total revenue of global online advertising market reached $42.8 billion in 2013, a growth of 17% over 2012.

As the result of the booming market, recent research focused on improving effectiveness in all market segments of online advertising. In the search engine market, for instance, a novel algorithm was proposed aiming to figure out the optimal allocation of online advertising space for “budget-constraint advertisers” (Mahdian, Nazerzadeh & Saberi, 2007). And in the online video game market, the influence that the match between games and
in-game ads could make on players’ purchase intention was studied (Chang, Yan, Zhang & Luo, 2010).

Among all market segments of online advertising, online video advertising market arouses our huge interest. It was estimated that the advertising investments in online videos in US would grow from $2.1 billion in 2009 to $4.3 billion in 2011, with the estimated CAGR of over 40% (Artero, 2010).

Despite the fast growth of the online video advertising market, numerous problems exist when we take into consideration the consumer experience of online video advertising. VanBoskirk et al. (2007) pointed out that 82% customers found the formats of online video advertising annoying. The main reason behind is that irrelevant ads are imposed on customers, which brings dramatic disturbance to the watching experience.

2. Existing RTB Model

![Figure 1. Existing RTB model, which consists of customer, publisher, SSP, DSP, Ad Exchange, advertiser and DMP.](image)

In the online video advertising market, a technology called real-time bidding (RTB) emerged in 2009 plays a significant role (Google, 2011). RTB was defined as the technology that leverages computer algorithms to intelligently bid for certain ads based on data about specific customers (Yuan, Wang & Zhao, 2013).

Money efficient in reaching target customers, RTB “has been the fastest growing segment of the online video business” (Forrester Consulting, 2013). According to the study of Forrester Consulting (2013), RTB spending grew at a CAGR of 57% between 2011 and 2014, achieving $1,141 million in 2014. Recent research shows the RTB technology has been continuously developed through various techniques. For instance, Chen, Berkhin, Anderson and Devanur (2011) maximized publishers’ revenue and achieved a good match between
advertisers’ campaigns and ad impressions with a novel algorithm.

Figure 1 shows the existing RTB model. We’ll take the online video advertising market as an example to explain the working mechanism of RTB. When a customer enters into a video platform such as YouTube, and chooses to watch a video, he or she has already triggered one round of RTB. On clicking the video, the ads supply and customer description are delivered to advertisers via publisher, Supply-Side Platforms (SSPs), Ad Exchange, Demand-Side Platforms (DSPs) and Data-Management Platforms (DMPs). Based on the customer description from DMPs which collect such data like browser cookies and purchase history, DSPs evaluate the value of this impression. The impression is the chance of showing ads to the specific customer. Having received bidding prices from DSPs, the Ad Exchange chooses the advertiser with the highest bidding price, and the winner’s advertisement is eventually presented to this customer (Zhang, Yuan & Wang, 2014).

Though mature in markets, there is still huge room for improvement within this model. The major formats of online video advertising involving RTB are: pre-rolls, post-rolls and overlay. Pre-rolls (post-rolls) are ads that are displayed before (after) the video content (VanBoskirk, Li, Katz & Lee, 2007). The time duration of pre-rolls and post-rolls usually vary from 10 to 60 seconds. Overlay ads appear when customers pause the video, and they are not dismissed until customers continue to watch the video. What’s more, customers are always treated as passive receptors of ads and speculated by advertisers through any hints from DMPs that may indirectly indicate their preferences.

3. Proposed RTB Model

![Proposed RTB model based on video processing technology, where automatic video processing and manual correction are applied and customer preferences are collected.](image)

International Conference on Innovation and Management, Sapporo, Japan, July 14-17, 2015.
Based on the observations above, we proposed a novel RTB model (Figure 2). In this model, when a customer clicks any object (commodity, celebrity, etc.) in the video and triggers RTB, specific information (properties, timestamp, location in a frame, etc.) about the object he or she clicks will be saved as customer preferences and sent to DMPs. Since in most cases, what customers are willing to click are what they desperately feel interested in, so the specific information helps advertisers to thoroughly understand the customer’s purchase preferences, identify and eventually reach their target customers. As for the customer, the most relevant ads are presented to him or her right after RTB ends (Zhang, Feng, Chang, Tan & Wu, 2015). The underlying technology is the automatic video processing and manual correction mechanism, which will be discussed thoroughly in the next section.

This new RTB model features extensive engagement and interactivity for customers, so customers wouldn’t have the feeling of being forced to watch irrelevant ads anymore. Ads tailored for customers are displayed only when customers truly want to watch them.

4. Proposed Video Processing Technology

To make the proposed RTB model technically feasible, we developed a novel video processing technology (Figure 3). This technology mainly includes one manual correction mechanism and four algorithms, which are algorithms of scene-splitting, object recognition, box selection and object tracking. Scene-splitting detects abrupt change of frames and lays a foundation for following procedures; recognition and box selection serve to identify the user-specified object from the video in first several frames of a scene; tracking algorithm captures the user-specified object in the remaining frames of a scene with the box selection result as an input; manual correction ensures the robustness and applicability of the technology in more complicated scenarios. Among them, we independently developed the manual correction mechanism and algorithms of scene-splitting and box selection, while algorithms of recognition and tracking were carefully chosen among dozens of state-of-art algorithms in computer vision field by their performance.

Figure 3. The video processing technology inputs videos, carries out scene splitting, object recognition, box selection object tracking, and outputs customer preferences.
So far this video processing technology has been validated in 15 various videos. For our research purpose, we assumed that customers only click cars. However, the success of this technology on dealing with cars proves the feasibility of the proposed model and paves the way to the success of handling other types of objects (Qin, Tan, Li, Zhang & Tang, 2014).

4.1 Scene-Splitting Algorithm

Firstly, we decomposed the entire video into numerous scenes. For the purpose of this paper, the scene here is defined to be a collection of consecutive frames where there should be no abrupt change of frame. The correlation coefficient (1) is used to decide whether there is an abrupt change between two frames,

\[
r = \frac{\sum_{m} \sum_{n} (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_{m} \sum_{n} (A_{mn} - \bar{A})^2\right)\left(\sum_{m} \sum_{n} (B_{mn} - \bar{B})^2\right)}}
\]

where \(A\) and \(B\) denote the image matrix of consecutive frames in a video. \(\bar{A}\) and \(\bar{B}\) are the mean of \(A\) and \(B\) respectively, and \(r\) denotes the correlation coefficient. To detect a scene changing using (1), we compare three values \(r_1, r_2\) and \(r_3\). We decide that a scene change happens only when \(r_1 - r_2 > \varepsilon\) and \(r_3 - r_2 > \varepsilon\), where \(\varepsilon\) is a predetermined threshold. After multiple experiments, we found that \(\varepsilon = 0.2\) is a suitable value.

The reason why we split the video into different scenes is to make sure that the following procedure of recognition and tracking can take place without abrupt changes, which enhances the accuracy and speed of video processing.

4.2 Object Recognition and Box Selection Algorithm

Secondly, we applied an OpenCV-based recognition algorithm called Latent SVM, which can also work on MATLAB (“Latent SVM”, n.d.). With a training set for object recognition, the recognition results are obtained by calling the Latent SVM detector in MATLAB. The recognition system is based on mixtures of deformable parts of objects, with a series of optimization (Felzenszwalb, Girshick, McAllester & Ramanan, 2010). This system relies heavily on new methods for discriminative training of classifiers that make use of latent information (Girshick, Felzenszwalb & McAllester, 2012). It also depends on efficient methods for matching deformable models to images (Girshick et al., 2012).

During validations, the success rate of the recognition algorithm in terms of capturing the target car is relatively high, but the problem is that many objects other than the target car are also captured. To address this problem, we developed a box selection algorithm, which aims to filter out the user-specified object from other recognition results. In the box selection algorithm, the HSV score is used for judgment, which is the sum of the difference between the color measured and the color the user defined. All boxes with negative scores are neglected, with the presupposition that Latent SVM works perfect. To get the HSV value of the whole box instead of pixels, the content of boxes are resized and converted from RGB to
HSV. The comparison terminates after the conversion of the last positive box, and the object with the lowest HSV score is selected as the recognition result.

4.3 Object Tracking Algorithm

Thirdly, we used the tracking algorithm of Online Robust Image Alignment (ORIA) (Wu, Shen & Ling, 2012). This tracking algorithm includes three key steps, which are batch image alignment, linearly reconstruction, and update of basis. Batch image alignment aligns all the images of the target objects to a fixed canonical template. Usually at the beginning of tracking, the first few target images are aligned to a template. Linearly reconstruction is used for the newly arriving images. Since we assumed that the aligned images are linearly correlated, the newly arriving image can be linearly reconstructed by the well-aligned image basis (Qin, Tan, Li, Zhang & Tang, 2014). Update of basis dynamically integrates the newly arrived image when it is different from all the images in the current basis (Wu et al., 2012).

The tracking algorithm plays an important role throughout the proposed technology because it computes most of the position information of objects.

4.4 Manual Correction and Graphical User Interface (GUI)

Lastly, we encapsulated all the algorithms with a graphical user interface (GUI) on MATLAB platform. By doing so, all four algorithms with one manual correction mechanism are integrated into the GUI (Figure 4). Since the resolution, scene-changing frequency and ambient disturbance can be significantly different in each video, the existence of the manual correction mechanism is critical in case the result of automatic video processing is far from satisfactory. Figure 5 shows the MATLAB GUI that is ready for use.

4.5 Customer Preferences

When a video is being processed, properties (object name, color, size, etc.), timestamp and location of the user-specified object are saved into a .json file and sent to DMPs as a reference for bidding.

Figure 4. Encapsulation of four algorithms with manual correction into the MATLAB GUI.
4.6 Live Demo

When video processing ends, the user-specified object is labeled with a red rectangle in frames where it appears. Demo of a processed video is available via Dropbox in case readers have further interest: https://www.dropbox.com/s/pknzwrofniuqflv/finish_720p.mp4?dl=0

5. Conclusion

To address the underlying issues of ineffectiveness and poor customer experience in the online video advertising industry, a novel RTB model is proposed. The RTB model highlights the concept of automatic video processing, manual correction and customer preferences.

In addition, a video processing technology was developed to technically support the proposed RTB model, which covers a manual correction mechanism and four algorithms including scene-splitting, object recognition, box selection and object tracking. Furthermore, a MATLAB GUI incorporating the manual correction mechanism with the existing algorithms fits the proposed model and technology in a real online video test. The technology was validated through satisfactory precision, accuracy, and processing speed in various 15 videos.

The proposed RTB model and the corresponding video processing technology create a significant customer engagement, providing a deeper and more specific insight. An advertiser can used a highly specific data to personalize the customer experience and expedite the purchase decision associated with the target customer.

Although the video processing technology has been validated, applying the proposed RTB model to a real business world is still challenging. This is because customers are not
used to click objects when watching videos. To change this situation and make full use of the proposed model, the industry needs to educate customers with a systematic process and teach them the interactive format of online video advertising. In the near future, customers will be instantly connected to what they want to watch simply by a mouse click, a finger touch or even an eye blink. Be prepared to usher in and embrace a new era of online advertising.

As for future work, validations of the video processing technology on different objects other than cars are recommended. In addition, since the proposed model hasn’t been put into the real market yet, the model is subject to minor or even major changes according to feedbacks from the market. The speed and accuracy of video processing should be further improved in order to meet the demanding needs of online advertising.

Acknowledgement

This research was supported by AVD Digital Media (AVD) and conducted at the University of Michigan – Shanghai Jiao Tong University Joint Institute (UM-SJTU). The authors thank Wayne Wu, Cayar Zhang and Andrew Ballen, who mentored, connected, and supported on AVD side. The authors thank Siyu Tang and Zhengrong Li from UM-SJTU Joint Institute for their contribution and efforts. The authors also thank Jason Chang who proofread and edited the manuscript.
References


Zhang, J., Feng, C., Chang, V., Tan., & Wu. W (2015, August). A novel business model based on real time bidding and online video interaction technology. Accepted for oral presentation in 2015 The European Business and Management Conference (Jul. 9 - 12)


